

MODÈLES PRÉDICTIFS (APPROCHE STATISTIQUE CLASSIQUE ET MACHINE LEARNING)

Durée & Format

Durée : 28 heures

Format disponible : Présentiel ou Distanciel

Tarifs

Tarif public intra : Consulter notre [site Internet](#)

Tarif public inter : Consulter notre [site Internet](#)

Public visé

Toute personne souhaitant comprendre la démarche de mise en place d'un modèle prédictif en combinant l'approche statistique et l'approche Machine Learning.

Prérequis

Il est indispensable que les participants aient :

- De bonnes connaissances sur les outils statistiques de base : statistiques descriptives, intervalles de confiance, p-value, risque alpha, tests d'hypothèses...
- Une connaissance de base de la régression est un atout.

Objectifs pédagogiques

A l'issue de cette formation, l'apprenant sera capable de :

- Définir le contexte et les objectifs des différentes méthodes de prédiction d'une variable quantitative
- Mettre en oeuvre et interpréter les résultats des méthodes suivantes
- Modèles Linéaires Généraux (Régressions classiques)
- Modèles Linéaires Généraux (Régression logistique, poisson, zip)
- Approche prédictive via les outils de Machine Learning (knn, svm, arbres de décisions, ...)
- Vérifier les conditions de mise en oeuvre des différentes méthodes
- Connaître les grandes lignes des concepts mathématiques inhérents à ces méthodes
- Identifier et prendre en compte les problèmes de multi colinéarités sur les modèles linéaires
- Mesurer la qualité d'ajustement et la qualité de prédiction (d'estimation des coefficients) du modèle
- Détecter et analyser les points aberrants / les points influents susceptibles d'altérer les modèles
- Utiliser les méthodes de type Machine Learning (knn, svm, arbres de décisions) dans le contexte de prédictions
- Connaître la démarche de validité d'un modèle (validation croisée, Loo, kfold,...)

Programme et déroulement

La formation en quelques mots

Cette formation s'adresse à des personnes souhaitant mettre en oeuvre des modèles prédictifs.

La particularité de cette formation réside dans le fait qu'elle va mixer deux approches :

- L'approche statistique classique basée sur les méthodes de régression type régression via les modèles linéaires généraux et généralisés.
- L'approche Machine Learning basée sur les méthodes d'apprentissage type knn,svm, Arbres de décisions, ...



L'intérêt de cette double approche est de sensibiliser les apprenants aux problématiques liées aux surajustement ou sous ajustements inhérents à ces méthodes. Il permet aussi d'effectuer parfois pour les apprenants une première approche dans le monde du Machine Learning, une des composantes utilisées en IA. La finalité des modèles prédictifs sera appliquée sur la prédiction de variables quantitatives mais aussi de variables qualitatives. L'idée est de passer en revue un certain nombre de méthodes permettant aux apprenants d'aller ensuite plus loin dans leur modélisation.

Deux compétences distinctes seront abordées lors de cette formation :

- **Comment fait-on pour prédire ?** Un passage en revue de différentes méthodes de prédictions sera réalisé (régression, régression logistique, régression de Poisson, knn, ...)
- **Comment s'assurer que l'on prédit bien ?** La mise en place d'une **stratégie** permettant de **valider** les prédictions effectuées (que l'on prédise des variables quantitatives ou qualitatives).

Jeux de données

Afin de s'approcher au mieux des réalités quotidiennes des praticiens, nous suggérons de nous appuyer pour l'animation pratique de thématiques et surtout de jeux de données reflétant le quotidien des apprenants.

Cet élément est un facteur de réussite pour la formation. Elle permet aux apprenants de :

- Se "reconnaître" dans les thèmes abordés,
- Mieux percevoir l'intérêt des notions étudiées
- S'approprier le contenu de la formation

Il sera donc pertinent que les apprenants puissent réfléchir en amont de la formation à des problématiques, jeux de données ou documents susceptibles d'être utilisés en support lors de la formation.

Outil logiciel

Cette formation n'est pas dédiée à la pratique d'un logiciel particulier. Les exercices et les illustrations se feront essentiellement autour de R.

Programme

Rappels sur les outils statistiques de base

- Grandeurs de position
- Grandeurs de dispersion
- Grandeurs de corrélation
- Intervalles de confiance
- Tests d'hypothèses (H_0 , p value)
- ...

Approche prédictive via les Modèles Linéaires Généraux

Régression simple et multiple

- Principes et concepts mathématiques inhérents à la régression linéaire simple
- Tests d'hypothèse de significativité du modèle
 - Utilisation du modèle Prédiction de valeurs individuelles
 - Intervalles de confiance des prédictions
- Traitement graphique des résultats
- Principes et concepts mathématiques inhérents à la régression linéaire multiple
- Inférence du modèle, inférence des variables (statistique de Fisher)
- Analyse des résidus
 - Calculs des résidus
 - Sens physique et statistique

- Homogénéité
- Distribution, Normalité
- Valeurs suspectes
- Analyses graphiques
- Valeurs suspectes et points influents
 - Résidus : résidus studentisés
 - Effet leviers
 - Distance de Cook

Le modèle de régression multiple

- De la régression simple à la régression multiple
- Significativité des coefficients
- Hiérarchie des coefficients
- Problèmes liés à la multi-colinéarité
- Mesures des colinéarités
 - Coefficient de corrélation
 - Vifs

Analyse des problèmes de multi colinéarités à travers la sélection des variables

- Détection de la multi colinéarité
 - Effets nocifs de la colinéarité entre variables explicatives
 - Outils de détection : corrélation, VIF, cohérence des signes des coefficients
- Traitement de la colinéarité
- Les solutions proposées
 - Expérimentations structurée (type matrices plans d'expériences)
 - Sélection de variables

Validation d'un modèle de régression

- Qualité du modèle
 - Qualité d'ajustement, coefficient de détermination R^2 , R^2 ajusté, R^2 prev
 - Qualité de prédiction, erreur d'estimation
- Utilisation du modèle
 - Prédiction (prévision) de valeurs individuelles
 - Intervalles de confiance des prédictions (prévisions)
- Illustration graphique des résultats
- Appliquer la validation croisée à un modèle de régression croisée
 - Méthode du leave one out
 - Méthode du k fold
 - Méthode de la séparation entre data d'entraînement et data de test

Approche prédictive via les Modèles Linéaires Généraux

Régression logistique

- Variable explicative et variable expliquée (continue / binaire)
- Différences entre la régression linéaire classique et la régression logistique
- Variables explicatives qualitatives, variables explicatives quantitatives
- Objectifs de la régression logistique
- Définition du modèle Logit (courbe sigmoïde)
- Conditions d'utilisation à respecter
- Estimation et interprétation des coefficients du modèle
- Test de significativité du modèle (validation du modèle)
- Tests d'apport d'une variable (test de Wald, tests sur les rapports de vraisemblance)

- Interprétation du Khi^2 de Wald
- Odds-ratios
- Parallèle odds ratios et risques relatifs
- Mise en oeuvre et analyse des résultats d'un modèle de régression logistique multiple
- Estimation et interprétation des coefficients du modèle multiple

Régression de Poisson & Régression Zip

- Le cadre des modèles linéaires Généralisés
- Différences entre modèles linéaires généraux et généralisés
- Identification d'une variable réponse Y à expliquer comme mesure de comptage
- Estimations des effets
- Lire et comprendre les coefficients
- Choix des variables
- Choix du modèle
- Diagnostic
- Problème de la surdispersion et des zéros
- ZIP, Zero Inflated Poisson
 - Variable de réponse comme mesure de comptage avec présence de zéro importante
 - Modélisation différenciée : $Y=0$ et $Y>0$
 - Lire et comprendre les coefficients
 - Choix des variables
 - Choix du modèle
 - Diagnostic
- Régression sur Y en proportion

Approches prédictives par les méthodes de type Machine Learning

Les approches Machine Learning pour les variables qualitatives

- La méthode knn
 - Principe de la méthode des plus proches voisins
 - Algorithme de calcul
 - Distances entre individus
 - Choix des proximités
 - Définition du paramètre k
- Ouverture sur les Svm
 - Démarche des svm
 - Notions de marge
 - Séparation linéaire
 - Séparation non linéaire
 - Fonction noyau
 - Réglage des hyper paramètres
- Autres approches...
 - Arbres de décisions
 - Random forest
 - ...
- Validations de méthodes prédictifs dans le contexte de l'approche Machine Learning

Certification

Consulter le programme sur notre [site internet](#) pour identifier si cette formation est certifiante.

Modalités pédagogiques

- Explications théoriques suivies de pratiques guidées puis de mises en autonomie.
- Exercices autonomes et réguliers pour assurer l'assimilation

Moyens et supports pédagogiques

Votre formation se déroule à distance avec :

- 1 ordinateur
- 1 vidéoprojecteur

Votre formation se déroule à distance avec :

- 1 ordinateur
- 1 connexion Internet
- 1 adresse e-mail valide
- 1 équipement audio (micro et enceintes ou casque)
- 1 Webcam (facultatif – dans l'idéal)
- 1 deuxième écran (facultatif – dans l'idéal)

Evaluation et suivi

Les objectifs pédagogiques sont évalués et suivis grâce à différentes **méthodes adaptées aux acquisitions de compétences visées**. Nos formateurs réalisent ces évaluations **tout au long de la formation**, que ce soit pendant les séances synchrones ou asynchrones. Voici une liste non exhaustive des méthodes d'évaluation pouvant être utilisées en formation :

- **Questionnaire** de vérification de connaissances (Quiz)
- Réalisation de **Travaux Pratiques** : production ou amélioration d'un fichier
- **Mises en situation** et grilles d'analyse
- **Serious Game** : jeu de rôles et analyse
- **Activités interactives** à travers l'utilisation d'une plate-forme connectée
- ...

Toutes nos formations intègrent **une auto-évaluation** via notre Extranet Stagiaire au début et à la fin de chaque action de formation. Cet outil offre à chacun la possibilité de mesurer sa progression par rapport aux objectifs pédagogiques visés et leurs atteintes.

Profil formateur

Nos formateurs sont certifiés à l'issue d'un parcours organisé par nos soins. Ils bénéficient d'un suivi de maintien et d'évolution de leurs compétences aussi bien au niveau technique que pédagogique. Chacun de nos formateurs a bénéficié d'une formation spécifique à l'animation de classe virtuelle et à l'utilisation des solutions de formation à distance du Groupe ARKESYS.

Support stagiaire

À l'issue de la formation, les exercices et travaux pratiques réalisés, leurs corrigés ainsi qu'un support de cours dématérialisé sera fourni à chaque stagiaire via son extranet stagiaire.

Accessible à tous

Cette formation est accessible à toute personne en situation de handicap. Notre référent handicap prendra contact avec les stagiaires concernés pour adapter l'animation à leurs besoins et rendre l'apprentissage accessible à tous. Enfin, nos centres de formation sont accessibles aux personnes à mobilité réduite. [En savoir plus](#)